

OIPE

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/09/801,852A**

DATE: 09/10/2001  
 TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt  
 Output Set: N:\CRF3\09102001\I801852A.raw

3 <110> APPLICANT: Bristol-Myers Squibb Company  
 4 Chiang, Shu-Jen  
 5 Jonathan, Basch  
 7 <120> TITLE OF INVENTION: DIRECT PRODUCTION OF DESACETYLCEPHALOSPORIN C  
 9 <130> FILE REFERENCE: ON0163  
 11 <140> CURRENT APPLICATION NUMBER: 09/801,852A  
 12 <141> CURRENT FILING DATE: 2001-03-08  
 14 <150> PRIOR APPLICATION NUMBER: 60/188,033  
 15 <151> PRIOR FILING DATE: 2000-03-09  
 17 <160> NUMBER OF SEQ ID NOS: 15  
 19 <170> SOFTWARE: PatentIn version 3.0  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 1716  
 23 <212> TYPE: DNA  
 24 <213> ORGANISM: Rhodosporidium toruloides  
 26 <400> SEQUENCE: 1

27 atgctcctta accttcac cctcgctcc ctcgctgcga cgctccagct cgccttgcc	60
29 tctccgacct ccctcgccg ccgcacgaaac ccaaacgagc cccctccgt cgtcgacctc	120
31 ggctacgccc gctaccaagg ctacttgaac gagaccgccc gactctactg gtggcgccga	180
33 atccgctacg cctcggtca gcgccttccag gctcctcaga cgcggcgac gcacaaggcc	240
35 gtccgcaacg cgactgagta tgaccgcata tggatggccg cttagcgaggg aaccaacacg	300
37 accaaggcgt tgccgcggcc tagcaacacgc tcgagcagcg cgccgcagaa acaggcgtcg	360
39 gaggattgcc ttttcctcaa ttcgttgc cccgcggct cgtgcgaggg cgacaatctt	420
41 cccgtcctcg ttacattca cgagggtggc tacgcctcg gcatgcgag caccggcagc	480
43 gactttgccg cttcaccaa gcacacggga accaagatgg tcgttgtaaa ttcgcgtac	540
45 cgtctcgga gctttggttt cctcgctggc caagccatga aggactacgg tgtaacgaac	600
47 gccggcttgc ttgaccagca attgcctt caatgggttc aacagcacgt ctgcgaagttc	660
49 ggcggcaacc ccatcacgt tacgatttg ggcgagtctg caggcgcagg gtccgttatg	720
51 aaccagatca ttgcgaacgg cgcaacacc gtcaggctc tcggctccaa gaaggccctc	780
53 ttccacgctg ccatcgctc ctccgttcc ctccccatcc aagccaagta caactcccc	840
55 ttccgcgacg tgcctactc ccaactcgctc tcggcgacaa actgcaccaa agccgcctcg	900
57 tccttcgtt gcctcgaaac ttcgtcgacgt gcccgcgtc ctgcggcgcc cgtgaagaac	960
59 tcggcgccgt tcccggtcg gttttggtc tatgtcccg tcgtcgacgg gaccccttg	1020
61 actgagcgcg cgtcgctct ttcgtccaa ggcaagaaga acctcaatgg caaccccttc	1080
63 accgggatca acaacctcgaa cgaaggattc atattcactg acgccactat tcagaacgac	1140
65 acgatcagcg accagtgcga gcgcgtctcc cagttcgacc gcctcctcg cggcccttc	1200
67 ccctacatca ctcggagga gcgcaggcc gtcgcaagc agtacccgat ctccgacgcg	1260
69 ccgtcaaaagg gcaacaccc ttctcgcatc tcggccgtca tcggcgactc gacccctcg	1320
71 tgcctcgaccc actggaccgc cgaggcgatc ggctcgatcc cccacaaggg cctctcgac	1380
73 tacgcgcgg ctcaccacgc gaccgacaac tcgtactaca tcggctccat ctggAACGGC	1440
75 aagaagtccg ttcgtccgtt ccagtccctt gacggcgccg tcggcggtt catcgagacg	1500
77 ttcaacccga acaacaacgc tggcaacaag accatcaacc ttactggcc gacgttcgac	1560
79 tcgggcaagc agctccctt caacacgcg acgaggaca ccctctctcc cggccgaccgc	1620
81 cgcacatcgat agacttcaag ctgaccgcac ttggcacga gccagaagac caagtgcgac	1680
83 ttctggcggt ggtcaatctc ggtgaacgcg ggtctc	1716

86 <210> SEQ ID NO: 2  
 87 <211> LENGTH: 572

**ENTERED**

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001  
TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt  
Output Set: N:\CRF3\09102001\I801852A.raw

88 <212> TYPE: PRT  
 89 <213> ORGANISM: Rhodosporidium toruloides  
 91 <400> SEQUENCE: 2  
 93 Met Leu Leu Asn Leu Phe Thr Leu Ala Ser Leu Ala Ala Thr Leu Gln  
 94 1 5 10 15  
 96 Leu Ala Phe Ala Ser Pro Thr Ser Leu Val Arg Arg Thr Asn Pro Asn  
 97 20 25 30  
 99 Glu Pro Pro Pro Val Val Asp Leu Gly Tyr Ala Arg Tyr Gln Gly Tyr  
 100 35 40 45  
 102 Leu Asn Glu Thr Ala Gly Leu Tyr Trp Trp Arg Gly Ile Arg Tyr Ala  
 103 50 55 60  
 105 Ser Ala Gln Arg Phe Gln Ala Pro Gln Thr Pro Ala Thr His Lys Ala  
 106 65 70 75 80  
 108 Val Arg Asn Ala Thr Glu Tyr Gly Pro Ile Cys Trp Pro Ala Ser Glu  
 109 85 90 95  
 111 Gly Thr Asn Thr Thr Lys Gly Leu Pro Pro Pro Ser Asn Ser Ser Ser  
 112 100 105 110  
 114 Ser Ala Pro Gln Lys Gln Ala Ser Glu Asp Cys Leu Phe Leu Asn Val  
 115 115 120 125  
 117 Val Ala Pro Ala Gly Ser Cys Glu Gly Asp Asn Leu Pro Val Leu Val  
 118 130 135 140  
 120 Tyr Ile His Gly Gly Tyr Ala Phe Gly Asp Ala Ser Thr Gly Ser  
 121 145 150 155 160  
 123 Asp Phe Ala Ala Phe Thr Lys His Thr Gly Thr Lys Met Val Val Val  
 124 165 170 175  
 126 Asn Leu Gln Tyr Arg Leu Gly Ser Phe Gly Phe Leu Ala Gly Gln Ala  
 127 180 185 190  
 129 Met Lys Asp Tyr Gly Val Thr Asn Ala Gly Leu Leu Asp Gln Gln Phe  
 130 195 200 205  
 132 Ala Leu Gln Trp Val Gln Gln His Val Ser Lys Phe Gly Gly Asn Pro  
 133 210 215 220  
 135 Asp His Val Thr Ile Trp Gly Glu Ser Ala Gly Ala Gly Ser Val Met  
 136 225 230 235 240  
 138 Asn Gln Ile Ile Ala Asn Gly Gly Asn Thr Val Lys Ala Leu Gly Leu  
 139 245 250 255  
 141 Lys Lys Pro Leu Phe His Ala Ala Ile Gly Ser Ser Val Phe Leu Pro  
 142 260 265 270  
 144 Tyr Gln Ala Lys Tyr Asn Ser Pro Phe Ala Glu Leu Leu Tyr Ser Gln  
 145 275 280 285  
 147 Leu Val Ser Ala Thr Asn Cys Thr Lys Ala Ala Ser Ser Phe Ala Cys  
 148 290 295 300  
 150 Leu Glu Ala Val Asp Ala Ala Leu Ala Ala Ala Gly Val Lys Asn  
 151 305 310 315 320  
 153 Ser Ala Ala Phe Pro Phe Gly Phe Trp Ser Tyr Val Pro Val Val Asp  
 154 325 330 335  
 156 Gly Thr Phe Leu Thr Glu Arg Ala Ser Leu Leu Leu Ala Lys Gly Lys  
 157 340 345 350  
 159 Lys Asn Leu Asn Gly Asn Leu Phe Thr Gly Ile Asn Asn Leu Asp Glu  
 160 355 360 365

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001  
TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt  
Output Set: N:\CRF3\09102001\I801852A.raw

162	Gly	Phe	Ile	Phe	Thr	Asp	Ala	Thr	Ile	Gln	Asn	Asp	Thr	Ile	Ser	Asp
163	370				375					380						
165	Gln	Ser	Gln	Arg	Val	Ser	Gln	Phe	Asp	Arg	Leu	Leu	Ala	Gly	Leu	Phe
166	385					390					395					400
168	Pro	Tyr	Ile	Thr	Ser	Glu	Glu	Arg	Gln	Ala	Val	Ala	Lys	Gln	Tyr	Pro
169						405				410						415
171	Ile	Ser	Asp	Ala	Pro	Ser	Lys	Gly	Asn	Thr	Phe	Ser	Arg	Ile	Ser	Ala
172						420				425						430
174	Val	Ile	Ala	Asp	Ser	Thr	Phe	Val	Cys	Pro	Thr	Tyr	Trp	Thr	Ala	Glu
175						435			440				445			
177	Ala	Phe	Gly	Ser	Ser	Ala	His	Lys	Gly	Leu	Phe	Asp	Tyr	Ala	Pro	Ala
178						450			455				460			
180	His	His	Ala	Thr	Asp	Asn	Ser	Tyr	Tyr	Ile	Gly	Ser	Ile	Trp	Asn	Gly
181	465					470				475						480
183	Lys	Lys	Ser	Val	Ser	Ser	Val	Gln	Ser	Phe	Asp	Gly	Ala	Leu	Gly	Gly
184						485				490						495
186	Phe	Ile	Glu	Thr	Phe	Asn	Pro	Asn	Asn	Asn	Ala	Ala	Asn	Lys	Thr	Ile
187						500			505							510
189	Asn	Pro	Tyr	Trp	Pro	Thr	Phe	Asp	Ser	Gly	Lys	Gln	Leu	Leu	Phe	Asn
190						515			520				525			
192	Thr	Thr	Thr	Arg	Asp	Thr	Leu	Ser	Pro	Ala	Asp	Pro	Arg	Ile	Val	Glu
193						530			535				540			
195	Thr	Ser	Ser	Leu	Thr	Asp	Phe	Gly	Thr	Ser	Gln	Lys	Thr	Lys	Cys	Asp
196						545			550			555				560
198	Phe	Trp	Arg	Gly	Ser	Ile	Ser	Val	Asn	Ala	Gly	Leu				
199						565				570						
201	<210>	SEQ ID NO:	3													
202	<211>	LENGTH:	2220													
203	<212>	TYPE:	DNA													
204	<213>	ORGANISM:	Rhodosporidium toruloides													
206	<400>	SEQUENCE:	3													
207	ggatccaccc	gaactctgtc	ccgctttctg	gctttcttcc	ttgctgtcgc	cccatcgct										60
209	tcccccactc	gccgccccatgc	tccttaaacct	cttcaccctc	gcctccctcg	ctgcgcacgct										120
211	ccagctcgcc	tttgccctc	cgacccctc	cgtccgcccgc	acgaacccaa	acgagcccc										180
213	tcccgtcgtc	gacctcggtc	acgccccgcta	ccaaggctac	ttgaacgaga	ccgccccgact										240
215	ctactgggtgg	cgcggaaatcc	gctacgcctc	ggctcagcgc	ttccaggctc	ctcagacgccc										300
217	cgcgcacgcac	aaggccgtcc	gcaacgcgac	tgagtatgg	ccgatctgtt	ggccggctag										360
219	cgagggaaacc	aacacgacca	agggcttggc	gccgccttagc	aacagctcg	gcagcgcgc										420
221	gcagaaaacag	gcgtcgagg	attgcctt	cctcaatgtc	gttgcccccg	ccggctcg										480
223	cgagggcgcac	aatctcccg	tcctcgatc	cattcacgga	gttgtggctacg	ccttcggcga										540
225	tgcgagcacc	ggcagcgact	ttgcggcctt	caccaaggcac	acgggaacca	agatggtcgt										600
227	tgttaaatctc	cagtaccgtc	tcggcagctt	tggttccctc	gttggccaag	ccatgaagga										660
229	ctacgggtgt	acgaacgccc	gcttgcttga	ccaggtgagt	ttcccgcatg	ataccggccc										720
231	acctttcgac	tcatgtgac	gcctctcccc	ctcgacgcaa	ttcgcccttc	aatgggttca										780
233	acagcacgtc	tcgaagttcg	gcccccaacc	cgatcacgtt	acgatttggg	gcgagtcgtc										840
235	aggcgcaggg	tccgttatga	accagatcat	tgcgaacgtg	agccacccga	accgatctcc										900
237	agccgacttt	ccccccccc	cccccccccgc	tgacctccct	cgtcttgcag	ggcggcaaca										960
239	ccgtcaaggc	tctcggtctc	aagaagcccc	tcttccacgc	tgccatcgcc	tcctccgtct										1020
241	tcctccctta	ccaagccaag	tacaactccc	cttcgccga	gtctcttac	tcccaactcg										1080

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001

TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw

243	tctcgccgac	aaactgcacc	aaagccgcct	cgtccttcgc	ttgcctcgaa	gctgtcgacg	1140
245	ctgcggcgct	cgctgcggcg	ggcgtgaaga	actcggcgcc	gttcccggtc	gggttttgtt	1200
247	ogtatgtccc	ggtcgtcgac	gggaccttct	tgactgagcg	cgcgtcgctc	cttctcgcca	1260
249	agggcaagaa	gaacctcaat	gcbcgtcggt	gcgagcttc	gagtgcgtca	ggatctcgct	1320
251	gacactgtcg	accggctcgc	agaaccttct	caccggatc	aacaacctcg	acgaagatga	1380
253	gttcccggtc	acggctctgt	tcgcccagcg	agactgactt	gttctttgc	gaagattacg	1440
255	attcatattc	actgacgcca	ctattcagaa	cgacacgatc	acgcgaccat	cgcagcgcgt	1500
257	ctcccagttc	gaccgcctcc	tcgcccggcct	cttcccctac	atcacctcgg	aggagcgcga	1560
259	ggccgtcgcg	aagcagtacc	cgatctccga	cgcgccgtca	aagggaacaaca	ccttctctcg	1620
261	catctcgcc	gtcatcgccg	actcgacctt	cgtgtcggtt	ccccgtcgtc	ttctccgagt	1680
263	atccgctga	cttcccgctt	gccccgacgt	gccccgaccta	ctggacccgc	gaggegttcg	1740
265	gctcgccgc	ccacaagggc	ctcttcgact	acgcgccgc	tcaccacgcg	accgacaact	1800
267	cgtactacat	cggctccatc	tggaacggca	agaagtcggt	ctcgccgtc	cagtccttcg	1860
269	acggcgcgct	cggcgcttc	atcgagacgt	tcaacccgaa	caacaacgct	gccaacaaga	1920
271	ccatcaaccc	ttactggccg	acggttcgact	cgggcaagca	gctcctcttc	aacacgacga	1980
273	cgagggacac	cctctctccc	gccgaccgc	gcatcggttga	gacttcaagc	ttgaccgact	2040
275	ttggcacgag	ccagaagacc	aagtgcgact	tctggcgtgg	gtcaatctcg	gtgaacgcgg	2100
277	gtctcttaggc	gtcttcctt	ccgacttcct	tcgttcttgc	gttgtttatt	cttgcagttc	2160
279	cgttgtatcg	gccattcgtg	cgtgtagctc	actcgagttat	agacgttggc	aagtgcgaaa	2220
282	<210>	SEQ ID NO:	4				
283	<211>	LENGTH:	544				
284	<212>	TYPE:	PRT				
285	<213>	ORGANISM:	Rhodosporidium toruloides				
287	<400>	SEQUENCE:	4				
289	Thr Asn Pro Asn Glu Pro Pro Val Val Asp Leu Gly Tyr Ala Arg						
290	1	5	10	15			
292	Tyr Gln Gly Tyr Leu Asn Glu Thr Ala Gly Leu Tyr Trp Trp Arg Gly						
293	20	25	30				
295	Ile Arg Tyr Ala Ser Ala Gln Arg Phe Gln Ala Pro Gln Thr Pro Ala						
296	35	40	45				
298	Thr His Lys Ala Val Arg Asn Ala Thr Glu Tyr Gly Pro Ile Cys Trp						
299	50	55	60				
301	Pro Ala Ser Glu Gly Thr Asn Thr Lys Gly Leu Pro Pro Pro Ser						
302	65	70	75	80			
304	Asn Ser Ser Ser Ala Pro Gln Lys Gln Ala Ser Glu Asp Cys Leu						
305	85	90	95				
307	Phe Leu Asn Val Val Ala Pro Ala Gly Ser Cys Glu Gly Asp Asn Leu						
308	100	105	110				
310	Pro Val Leu Val Tyr Ile His Gly Gly Gly Tyr Ala Phe Gly Asp Ala						
311	115	120	125				
313	Ser Thr Gly Ser Asp Phe Ala Ala Phe Thr Lys His Thr Gly Thr Lys						
314	130	135	140				
316	Met Val Val Val Asn Leu Gln Tyr Arg Leu Gly Ser Phe Gly Phe Leu						
317	145	150	155	160			
319	Ala Gly Gln Ala Met Lys Asp Tyr Gly Val Thr Asn Ala Gly Leu Leu						
320	165	170	175				
322	Asp Gln Gln Phe Ala Leu Gln Trp Val Gln Gln His Val Ser Lys Phe						
323	180	185	190				
325	Gly Gly Asn Pro Asp His Val Thr Ile Trp Gly Glu Ser Ala Gly Ala						

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001  
TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt  
Output Set: N:\CRF3\09102001\I801852A.raw

326	195	200	205
328	Gly Ser Val Met Asn Gln Ile Ile Ala Asn Gly Gly Asn Thr Val Lys		
329	210	215	220
331	Ala Leu Gly Leu Lys Lys Pro Leu Phe His Ala Ala Ile Gly Ser Ser		
332	225	230	235
334	240		
335	Val Phe Leu Pro Tyr Gln Ala Lys Tyr Asn Ser Pro Phe Ala Glu Leu		
337	245	250	255
338	Leu Tyr Ser Gln Leu Val Ser Ala Thr Asn Cys Thr Lys Ala Ala Ser		
340	260	265	270
341	Ser Phe Ala Cys Leu Glu Ala Val Asp Ala Ala Ala Leu Ala Ala Ala		
343	275	280	285
344	Gly Val Lys Asn Ser Ala Ala Phe Pro Phe Gly Phe Trp Ser Tyr Val		
346	290	295	300
347	Pro Val Val Asp Gly Thr Phe Leu Thr Glu Arg Ala Ser Leu Leu Leu		
350	305	310	315
352	320		
353	Ala Lys Gly Lys Lys Asn Leu Asn Gly Asn Leu Phe Thr Gly Ile Asn		
355	325	330	335
356	340	345	350
358	Asn Leu Asp Glu Gly Phe Ile Phe Thr Asp Ala Thr Ile Gln Asn Asp		
361	355	360	365
364	Thr Ile Ser Asp Gln Ser Gln Arg Val Ser Gln Phe Asp Arg Leu Leu		
365	370	375	380
367	385	390	395
368	Arg Ile Ser Ala Val Ile Ala Asp Ser Thr Phe Val Cys Pro Thr Tyr		
370	405	410	415
371	Trp Thr Ala Glu Ala Phe Gly Ser Ser Ala His Lys Gly Leu Phe Asp		
373	420	425	430
374	373	375	380
376	Tyr Ala Pro Ala His His Ala Thr Asp Asn Ser Tyr Tyr Ile Gly Ser		
377	435	440	445
379	450	455	460
380	Ala Leu Gly Gly Phe Ile Glu Thr Phe Asn Pro Asn Asn Asn Ala Ala		
382	465	470	475
383	480		
385	Asn Lys Thr Ile Asn Pro Tyr Trp Pro Thr Phe Asp Ser Gly Lys Gln		
386	485	490	495
388	Leu Leu Phe Asn Thr Thr Arg Asp Thr Leu Ser Pro Ala Asp Pro		
389	500	505	510
391	515	520	525
392	Arg Ile Val Glu Thr Ser Ser Leu Thr Asp Phe Gly Thr Ser Gln Lys		
393	530	535	540
394	<210> SEQ ID NO: 5		
395	<211> LENGTH: 15		
396	<212> TYPE: PRT		
397	<213> ORGANISM: Rhodosporidium toruloides		
398	<400> SEQUENCE: 5		
399	Thr Asn Pro Asn Glu Pro Pro Val Val Asp Leu Gly Tyr Ala		

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001

TIME: 11:23:04

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw